Cygnus A super-resolved via convex optimisation from VLA data

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **476,** 2853–2866 (2018) doi:10.1093/mnras/sty372
Advance Access publication 2018 February 13

Cygnus A super-resolved via convex optimization from VLA data

A. Dabbech,[1]★ A. Onose,[1] A. Abdulaziz,[1] R. A. Perley,[2] O. M. Smirnov[3,4] and Y. Wiaux[1]

[1]*Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK*
[2]*National Radio Astronomy Observatory, PO Box 0, Soccoro, NM 87801, USA*
[3]*Department of Physics and Electronics, Rhodes University, PO Box 94, Grahamstown 6140, South Africa*
[4]*SKA South Africa, 3rd Floor, The Park, Park Road, Pinelands 7405, South Africa*

Accepted 2018 January 27. Received 2017 December 20; in original form 2017 October 24

ABSTRACT

We leverage the Sparsity Averaging Re-weighted Analysis approach for interferometric imaging, that is based on convex optimization, for the super-resolution of Cyg A from observations at the frequencies 8.422 and 6.678 GHz with the Karl G. Jansky Very Large Array (VLA). The associated average sparsity and positivity priors enable image reconstruction beyond instrumental resolution. An adaptive Preconditioned primal-dual algorithmic structure is developed for imaging in the presence of unknown noise levels and calibration errors. We demonstrate the superior performance of the algorithm with respect to the conventional CLEAN-based methods, reflected in super-resolved images with high fidelity. The high-resolution features of the recovered images are validated by referring to maps of Cyg A at higher frequencies, more precisely 17.324 and 14.252 GHz. We also confirm the recent discovery of a radio transient in Cyg A, revealed in the recovered images of the investigated data sets. Our MATLAB code is available online on GitHub.

Key words: techniques: image processing – techniques: interferometric – galaxies: individual: Cygnus A.

1 INTRODUCTION

New imaging techniques and algorithmic structures for radio interferometry have been extensively investigated in the recent years. The main objectives are: first to meet the next-generation instruments' capabilities in producing maps of the radio sky with unprecedented depth and resolution and secondly to cope with the sheer volume of the acquired data. Recently proposed compressive sensing techniques using convex optimization for radio interferometric (RI) imaging have been shown to be very promising, potentially superseding the standard CLEAN-based techniques in terms of quality (e.g. Högbom 1974; Clark 1980; Schwab & Cotton 1983; Wakker & Schwarz 1988; Bhatnagar & Cornwell 2004; Cornwell, Golap & Bhatnagar 2008), while, in principle, showing scalability to big data. The general approach consists in minimizing a sum of convex functions. These include data fidelity terms and relevant regularizations for RI images, such as sparsity and positivity (typically for an intensity image). In particular, Carrillo, McEwen & Wiaux (2012) proposed the Sparsity Averaging Re-weighted Analysis (SARA) approach, where sparsity-by-analysis of the sky estimate is promoted in a collection of bases by solving consecutive re-weighted ℓ_1 problems. Several algorithms based on convex optimization have been proposed to solve the SARA minimization problem, such as Douglas-Rachford splitting algorithm (Carrillo et al. 2012) and the Simultaneous Direction Method of Multipliers (SDMM; Carrillo, McEwen & Wiaux 2014). Onose et al. (2016) have proposed a primal-dual (PD) algorithmic structure for RI imaging, in the context of which full splitting of the different functions involved is achieved, resulting in a highly parallelizable algorithm.

More recently, Onose, Dabbech & Wiaux (2017) have proposed an accelerated PD algorithmic structure reconciling two common data weighting schemes in RI imaging, namely natural and uniform weighting. On the one hand, natural weighting, by accounting only for the noise statistics, provides optimal sensitivity. On the other hand, uniform weighting, by additionally incorporating the density of the sampling, modifies the effective sampling and consequently, the associated point spread function (PSF), with the aim of optimizing the resolution achieved within a finite number of iterations. Yet, this scheme reduces the overall sensitivity, since the scarcely sampled–thus noisy–measurements at the high spatial frequencies are overweighted, while the highly sampled–thus sensitive–measurements at the low spatial frequencies are downweighted. In the context of the PD algorithm, the effect of the density of the sampling is instead cast in terms of the convergence speed. The algorithmic structure proposed in Onose et al. (2017), dubbed preconditioned primal-dual (PPD) enforces data fidelity on the naturally weighted data via non-Euclidean proximity operators, where projections on to the ℓ_2 balls are generalized as projections on to ℓ_2 ellipsoids incorporating the uniform weights. These

★ E-mail: a.dabbech@hw.ac.uk

ellipsoid projections provide accelerated convergence, thereby enabling simultaneous optimization of the dynamic range and resolution of the recovered image. In summary, the versatility of convex optimization with respect to the choice of both the regularization priors (e.g. SARA) and the algorithmic structure (e.g. PPD) has opened the door to tremendous potential for enhancement of the quality of RI imaging when compared to CLEAN-based techniques.

In this paper, we leverage the PPD algorithmic structure for solving the SARA minimization problem and provide high-fidelity, high-resolution imaging of Cyg A from Karl G. Jansky Very Large Array (VLA) observations at the frequencies 8.422 and 6.678 GHz. In addition to the thermal noise, the two data sets are corrupted with multiplicative calibration errors, which are likely due to the antennas' pointing errors. These are particularly important for strong radio sources such as Cyg A and indeed tend to dominate the thermal noise. We propose an *adaptive* version of the PPD algorithmic structure aiming for the estimation of the unknown levels of the noise and calibration errors in the data. On the one hand, high-fidelity maps of Cyg A at both frequencies are achieved thanks to the accurate estimation of the ℓ_2 constraints on the data. On the other hand, super-resolved representations of Cyg A are obtained thanks to the average sparsity-promoting and positivity priors of SARA. Comparison with the standard Multi-Scale CLEAN algorithm (MS-CLEAN) (Cornwell 2008) indicates the superior performance of the adaptive PPD algorithm. Furthermore, the recent discovery of a faint transient radio source in the inner core of Cyg A, reported in Perley et al. (2017), is confirmed at the reconstructed maps of adaptive PPD.

The remainder of the article is structured as follows. In Section 2, we revisit the RI inverse imaging problem and the compressive sensing-based image reconstruction approach SARA. In addition, we briefly describe the PPD algorithmic structure solving for SARA. In Section 3, we present the adaptive PPD algorithm. Cyg A imaging results from two data sets are presented in Section 4. The achieved super-resolution with adaptive PPD is analysed in Section 5. We also report the detection of a secondary black hole in the super-resolved images of adaptive PPD. Finally, conclusions are stated in Section 6.

2 SPARSE RI IMAGING TO DATE

In this section, we revisit the RI measurement model and the minimization problem to recover radio images that is based on sparse representations and denoted SARA. We also review the PD algorithmic structure solving the SARA minimization problem, recently proposed in Onose et al. (2017).

2.1 RI imaging problem

RI data are Fourier measurements of the sky intensity modulated with the so-called Direction Dependent Effects (DDEs).[1] These include the primary beam of the instrument and distortions induced by the propagation medium. Let $(\boldsymbol{u}, w)$ be the components of a baseline in units of the wavelength, with w being the coordinate in the direction of the line of sight and $\boldsymbol{u} = (u, v)$ lying on its perpendicular plane. Assuming a monochromatic and non-polarized radiation, an RI measurement $V(\boldsymbol{u}, w)$ reads

$$V(\boldsymbol{u}, w) = \int G(\boldsymbol{l}, w) I(\boldsymbol{l}) \mathrm{e}^{-2i\pi \boldsymbol{u}\cdot\boldsymbol{l}} \mathrm{d}^2\boldsymbol{l}, \tag{1}$$

where $\boldsymbol{l}$ are the coordinates of a radio point source in the plane tangent to the celestial sphere and $I(\boldsymbol{l})$ is the unknown sky surface brightness at the position $\boldsymbol{l}$. $G(\boldsymbol{l}, w)$ stands for the DDEs, including the w-modulation that is resulting from the non-coplanarity of the radio interferometer and is given by $c(w, \boldsymbol{l}) = \mathrm{e}^{-2i\pi w(\sqrt{1-|\boldsymbol{l}|^2}-1)}$. The problem of recovering the sky intensity image from the radio measurements is an inverse problem and its discretized version reads

$$\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{n}, \text{ with } \boldsymbol{\Phi} = \boldsymbol{\Theta G F Z S}, \tag{2}$$

where $\boldsymbol{x} \in \mathbb{R}_+^N$ is the intensity image of interest and $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ is the mapping operator from the image domain to the visibility space. The data $\boldsymbol{y}$ are the naturally weighted RI measurements i.e. $\boldsymbol{y} = \boldsymbol{\Theta}\tilde{\boldsymbol{y}}$, where $\tilde{\boldsymbol{y}} \in \mathbb{C}^M$ are the RI measurements and the operator $\boldsymbol{\Theta} \in \mathbb{R}_+^{M\times M}$ accounts for the noise statistics and is a diagonal matrix whose elements are the square root of the natural weights. The operator $\mathbf{G} \in \mathbb{C}^{M\times oN}$ is the so-called gridding matrix, interpolating the RI measurements from the discrete Fourier components of the sky $\boldsymbol{x}$ that are lying on a regular grid. Its rows are convolutional kernels, each centred at the corresponding spatial frequency $\boldsymbol{u}_{\ell\in\{1,\ldots,M\}}$. $\mathbf{F} \in \mathbb{C}^{oN\times oN}$ is the Fourier matrix. $\mathbf{Z}$ is a zero-padding operator in the image space, allowing for a fine Fourier grid, thus a more accurate interpolation. $\mathbf{S} \in \mathbb{R}^{N\times N}$ is correcting for the convolution in the Fourier domain through $\mathbf{G}$. Note that, for the data investigated herein, the probed field of view is narrow (i.e. $\|\boldsymbol{l}\|_2 \ll 1$). Thus, the w-modulation reduces to a flat function $c(w, \boldsymbol{l}) = 1, \forall \boldsymbol{l}$. In general, it can be efficiently incorporated in the operator $\mathbf{G}$ as measurement-dependent convolutional kernels (Dabbech et al. 2017).

2.2 Sparse image reconstruction approach

Due to the incompleteness of the Fourier sampling and the presence of the noise, the problem of recovering the image of the sky $\boldsymbol{x}$ from the noisy measurements $\boldsymbol{y}$ is ill-posed. In order to reconstruct a reliable approximation, prior knowledge on the unknown sky is crucial and has to be considered in the imaging problem. In particular, the sparsity of the signal in adequate data representation spaces has been extensively adopted for RI imaging in the recent years (e.g. Wiaux et al. 2009b; Li, Cornwell & de Hoog 2011; Carrillo, McEwen & Wiaux 2012; Dabbech, Mary & Ferrari 2012; Dabbech et al. 2015; Garsden et al. 2015). Sparse regularizations are backed by the theory of compressive sensing (Candès 2006). The theory proves that an exact recovery of the unknown signal can be achieved from noisy and incomplete measurements provided that the sensing basis $\boldsymbol{\Phi}$ is incoherent with the sparsity basis $\boldsymbol{\Psi}$ of the signal. Moreover, these regularizations can be easily enforced via convex functions. The resulting imaging problem can be efficiently solved using convex optimization.

In the present work, we adopt the following minimization problem, named the SARA approach, solving the inverse problem set in (2) and originally proposed in Carrillo et al. (2012)

$$\min_{\boldsymbol{x}} \| \mathbf{W}\boldsymbol{\Psi}^\dagger \boldsymbol{x} \|_1 \text{ s.t.} \begin{cases} \| \boldsymbol{y} - \boldsymbol{\Phi x} \|_2 \leq \epsilon, \\ \boldsymbol{x} \geq \mathbf{0}, \end{cases} \tag{3}$$

where ϵ is the ℓ_2 norm of the noise and constitutes the bound on the data fidelity term. Sparsity of the unknown signal $\boldsymbol{x}$ is promoted by analysis, i.e. its projection in a redundant data representation space

[1] A particular case of DDEs are the so-called direction-independent effects (DIEs). These are constant complex-valued modulations in the image domain.

is sparse. The adopted sparsity basis $\boldsymbol{\Psi} = [\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_b]$ is a collection of nine orthogonal bases: the Dirac basis and the eight first Daubechies wavelet bases. The most intuitive measure of sparsity is the ℓ_0 norm. However, being non-convex and yielding NP-hard problems, it is often replaced by its convex relaxation the ℓ_1 norm. In SARA, a re-weighted ℓ_1 norm is adopted, where the weighting matrix is $\mathbf{W} = [\mathbf{W}_1, \ldots, \mathbf{W}_b]$, $\forall i \in \{1, \ldots, b\}$, $\mathbf{W}_i \in \mathbb{R}_+^{N\times N}$ being diagonal matrices. Solving a sequence of re-weighted ℓ_1 minimization problems leads to nearly strictly sparse signals (Candès, Wakin & Boyd 2008). The SARA approach has been shown to provide superior imaging quality to CLEAN-based approaches both on simulations and few real data sets (Carrillo et al. 2014; Onose et al. 2016, 2017; Pratley et al. 2017).

Data fidelity can be enforced in a distributed manner by splitting the data and the measurement operator into d blocks as described in Carrillo et al. (2014), Onose et al. (2016), Onose et al. (2017). In this setting, the minimization task (3) equivalently reads

$$\min_{\boldsymbol{x}} \| \mathbf{W}\boldsymbol{\Psi}^\dagger \boldsymbol{x} \|_1 \text{ s.t.} \begin{cases} \| \boldsymbol{y}_j - \boldsymbol{\Phi}_j \boldsymbol{x} \|_2 \leq \epsilon_j, \ \forall j \in \{1, \ldots, d\}, \\ \boldsymbol{x} \geq \mathbf{0}, \end{cases} \quad (4)$$

where for each $j \in \{1, \ldots, d\}$, $\boldsymbol{\Phi}_j = \boldsymbol{\Theta}_j \mathbf{G}_j \mathbf{FZS}$ is the measurement operator associated with the data block $\boldsymbol{y}_j \in \mathbb{C}^{M_j}$. ϵ_j is the ℓ_2 norm of the noise $\boldsymbol{n}_j \in \mathbb{C}^{M_j}$ and consequently the ℓ_2 bound on the data block fidelity constraint. Note that the constraint formulation of the minimization problem (4) (and its equivalent formulation (3)) assumes accurate knowledge of the noise. This is challenging in real applications, in particular in the presence of significant calibration errors. These imply that the models of the operators $\boldsymbol{\Phi}_j, j \in \{1, \ldots, d\}$, are approximate. The ℓ_2 bounds $\{\epsilon_j\}_{j\in\{1,\ldots,d\}}$ will therefore have to account not only for the thermal noise but also the calibration errors.

Different algorithmic structures based on convex optimization have been adopted to address the minimization problem (4). These solvers fit within the proximal splitting methods (see Combettes & Pesquet 2011, for a review). In this framework, a minimization task is solved iteratively with each function handled individually. Typically, the differentiable functions are minimized using their gradient and the non-smooth functions are solved via their proximity operators. Carrillo et al. (2014) adopted the SDMM. The algorithm involves matrix inversions on the updates of the estimates of the solution. This results in a computational bottleneck when recovering large-sized images, despite the separability of the different functions involved in the minimization task. Onose et al. (2016) proposed two algorithmic structures showing high scalability to big data. These are a sub-iterative variant of the Alternating Direction Method of Multipliers (ADMM) and the PD algorithm using forward–backward iterations. On the one hand, the ADMM-based algorithmic structure presents a partial splitting of the functions involved. On the other hand, PD allows for a full splitting of all the operators and functions. Furthermore, at each iteration, randomized updates on the different variables involved are allowed. The computational load per iteration is therefore reduced. The algorithm is also shipped with a preconditioning functionality bringing accelerated convergence (Onose et al. 2017), hence the greater scalability of PD to big data.

2.3 The PD algorithmic structure

In the PD algorithm, the following primal problem is solved

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \gamma \sum_{i=1}^{b} l(\mathbf{W}_i \boldsymbol{\Psi}_i^\dagger \boldsymbol{x}) + \sum_{j=1}^{d} h_j(\boldsymbol{\Phi}_j \boldsymbol{x}), \quad (5)$$

together with its dual formulation

$$\min_{\substack{\boldsymbol{u}_i \\ \boldsymbol{v}_j}} f^* \left(- \sum_{i=1}^{b} \boldsymbol{\Psi}_i \mathbf{W}_i \boldsymbol{u}_i - \sum_{j=1}^{d} \boldsymbol{\Phi}_j^\dagger \boldsymbol{v}_j \right) + \frac{1}{\gamma} \sum_{i=1}^{b} l^*(\boldsymbol{u}_i) + \sum_{j=1}^{d} h_j^*(\boldsymbol{v}_j). \quad (6)$$

The parameter γ is free and only affects the convergence speed, $\boldsymbol{x}$ is the primal variable that is the unknown image of the sky, and $\boldsymbol{u}_{i\in\{1,\ldots,b\}}$, $\boldsymbol{v}_{j\in\{1,\ldots,d\}}$ are the dual variables associated with the sparsity priors and the data fidelity terms, respectively. The notation $(^*)$ stands for the Legendre-Fenchel conjugate function. Note that, in the formulation (5), further splitting of the sparsity prior with respect to each sparsity basis is achieved thanks to the separability of the ℓ_1 norm. Moreover, the constraints are reformulated using the indicator function.[2] The functions involved are $f = \iota_{\mathbb{R}_+^N}$, enforcing the positivity and the reality of the unknown signal, $l = \|.\|_1$, imposing sparsity by analysis of the signal in the basis $\boldsymbol{\Psi}_i$, $\forall i \in \{1, \ldots, b\}$ and $h_j = \iota_{\mathcal{B}_j}$, where $\mathcal{B}_j = \{\boldsymbol{z} \in \mathbb{C}^{M_j} : \|\boldsymbol{z} - \boldsymbol{y}_j\|_2 \leq \epsilon_j\}$, are the data fidelity terms, enforcing the residual data blocks to be within the ℓ_2 balls $\mathcal{B}_j$, $\forall j \in \{1, \ldots, d\}$. The formulated problem (5) is analogous to the problem (4).

The different functions involved in (5) and (6) are non-differentiable; therefore, they are minimized via their proximity operators. Considering a lower semicontinuous and proper convex function g, its proximal operator is defined as

$$(\forall z), \ \text{prox}_g(z) \triangleq \underset{\bar{z}}{\text{argmin}}\, g(\bar{z}) + \frac{1}{2}\|z - \bar{z}\|_2^2. \quad (8)$$

Following this definition, the function f is minimized via projections on the positive and real orthant, the sparsity function l is minimized via soft-thresholding operators, and the data fidelity terms h_j are minimized via projections on to the ℓ_2 balls $\mathcal{B}_j$, simultaneously. The proximal operators of the conjugate functions involved in problem (6) are obtained from those of problem (5) by the Moreau decomposition, as follows:

$$(\forall z), \ \text{prox}_{g^*}(z) \triangleq z - \text{prox}_g(z). \quad (9)$$

The two minimization tasks (5) and (6) are solved via forward–backward steps updating the dual and the primal variables (Pesquet & Repetti 2015). These consist in a gradient descent step coupled with a proximal update. In analogy with CLEAN (e.g. Cotton–Schwab CLEAN; Schwab & Cotton 1983), the algorithm can be understood as being composed of complex CLEAN-like steps performed in parallel in data, prior and image spaces (the reader is directed to Onose et al. 2016, for further details).

2.4 The PPD algorithmic structure

Onose et al. (2017) have recently proposed the algorithmic structure PPD, where an acceleration strategy within the PD algorithm is adopted. It consists of incorporating a priori knowledge on the data when enforcing fidelity to the naturally weighted data. This is made feasible thanks to the generalized definition of the proximal operator

[2] Considering a convex set $\mathcal{C}$, its indicator function is defined as

$$(\forall z), \ \iota_{\mathcal{C}}(z) \triangleq \begin{cases} 0 & z \in \mathcal{C} \\ +\infty & z \notin \mathcal{C}. \end{cases} \quad (7)$$

Algorithm 1 Adaptive preconditioned forward–backward PD.

1: **given** $\boldsymbol{x}^{(0)}, \bar{\boldsymbol{x}}^{(0)}, \boldsymbol{u}_i^{(0)}, \boldsymbol{v}_j^{(0)}, \mathbf{W}_i, \mathbf{U}_j, \epsilon_j^{(0)}, \kappa, \tau, \eta, \zeta, \boldsymbol{\gamma}, P, p_j^{(0)}, \sigma^{(0)}$
2: **repeatfor** $t = 1, \ldots$
3: **run simultaneously**
4: $\forall j \in \{1, \ldots, d\}$ **do in parallel**
5: $\boldsymbol{v}_j^{(t)} = \mathbf{U}_j^{1/2}\left(\mathcal{I} - \mathcal{P}_{\mathcal{E}_j^{(t-1)}}\right)\left(\mathbf{U}_j^{-1/2}\boldsymbol{v}_j^{(t-1)} + \mathbf{U}_j{}^{-1/2}\boldsymbol{\Phi}_j\bar{\boldsymbol{x}}^{(t-1)}\right)$
6: $\mu_j^{(t)} = \|\boldsymbol{y}_j - \boldsymbol{\Phi}_j\boldsymbol{x}^{(t-1)}\|_2$
7: **if** $\sigma^{(t-1)} < \gamma_1$ **and** $t - p_j^{(t-1)} \geq P$ **and** $\frac{|\mu_j^{(t)} - \epsilon_j^{(t-1)}|}{\epsilon_j^{(t-1)}} > \gamma_2$
8: $\epsilon_j^{(t)} = \gamma_3\mu_j^{(t)} + (1-\gamma_3)\epsilon_j^{(t-1)}$
9: $p_j^{(t)} = t$
10: **else set** $\epsilon_j^{(t)} = \epsilon_j^{(t-1)}$ **and** $p_j^{(t)} = p_j^{(t-1)}$
11: **end**
12: $\forall i \in \{1, \ldots, b\}$ **do in parallel**
13: $u_i^{(t)} = \left(\mathcal{I} - \mathcal{S}_{\kappa\|\boldsymbol{\Psi}\boldsymbol{W}\|_S}\right)\left(\boldsymbol{u}_i^{(t-1)} + \boldsymbol{W}_i^\dagger\boldsymbol{\Psi}_i^\dagger\bar{\boldsymbol{x}}^{(t-1)}\right)$
14: **end**
15: **end**
16: $\boldsymbol{x}^{(t)} = \mathcal{P}_{\mathbb{R}_+^N}\left(\boldsymbol{x}^{(t-1)} - \tau\left(\eta\sum_{j=1}^{d}\boldsymbol{\Phi}_j^\dagger v_j^{(t)} + \zeta\sum_{i=1}^{b}\boldsymbol{\Psi}_i\boldsymbol{W}_i\boldsymbol{u}_i^{(t)}\right)\right)$
17: $\bar{\boldsymbol{x}}^{(t)} = 2\boldsymbol{x}^{(t)} - \boldsymbol{x}^{(t-1)}$
18: $\sigma^{(t)} = \frac{\|\boldsymbol{x}^{(t-1)} - \boldsymbol{x}^{(t)}\|_2}{\|\boldsymbol{x}^{(t)}\|_2}$
19: **until convergence**
20: **output** $\boldsymbol{x}^{(t)}, \bar{\boldsymbol{x}}^{(t)}, \boldsymbol{u}_i^{(t)}, \boldsymbol{v}_j^{(t)}$

which, considering a strongly positive self-adjoint linear operator $\mathbf{U}$, reads

$$(\forall z),\ \operatorname{prox}_g^{\mathbf{U}}(z) \triangleq \underset{\bar{z}}{\operatorname{argmin}}\, g(\bar{z}) + \frac{1}{2}(z-\bar{z})^\dagger\mathbf{U}(z-\bar{z}). \tag{10}$$

Following this definition, for each data block indexed by $j \in \{1, \ldots, d\}$, the ℓ_2 projections are performed on skewed balls. Conceptually, these are equivalent to projections on to the ellipsoids $\mathcal{E}_j$, defined by $\mathcal{E}_j = \{\bar{\boldsymbol{s}} \in \mathbb{C}^{M_j} : \|\mathbf{U}_j^{-\frac{1}{2}}\bar{\boldsymbol{s}} - \boldsymbol{y}_j\|_2 \leq \epsilon_j\}$, which are then moved to the ℓ_2 balls $\mathcal{B}_j$ via the linear operator $\mathbf{U}_j^{-1/2}$ (see Onose et al. 2017, for further details). In this setting, the operator $\mathbf{U}$ incorporates the prior information on the data and acts as a preconditioning matrix affecting only the speed of convergence while enforcing the data fidelity with respect to the naturally weighted data. A relevant choice of the preconditioning matrix $\mathbf{U}$ to ensure a faster convergence involves the uniform weights (Onose et al. 2017). More precisely, the matrix $\mathbf{U} \in \mathbb{R}_+^{M\times M}$ is set as a diagonal matrix, whose elements are inversely proportional to the density of the sampling at the vicinity of the probed Fourier modes. The more non-uniform the original density of the sampling over the Fourier plane, the more effective the approach.

The PPD algorithmic structure is given in Algorithm 1. Note that steps are specific to adaptive PPD, proposed in the following section. All the dual variables involved in the problem (6) are updated in parallel via forward–backward steps. The dual variables associated with the data terms $\boldsymbol{v}_{j\in\{1,..,d\}}$ are updated in Step 5, where projections on the ellipsoids $\mathcal{E}_j$ are performed and the dual variables associated with the sparsity prior $\boldsymbol{u}_{i\in\{1,..,b\}}$ are updated via soft-thresholding operations controlled by the parameter κ in Step 13. These dual variables are then utilized in Step 17, as incremental variables in the update of the primal variable $\boldsymbol{x}$, i.e. the image of interest. The latter is followed with a projection on to the real positive orthant.

Algorithm 2 Re-weighting scheme.

1: **given** $\boldsymbol{x}^{(0)}, \bar{\boldsymbol{x}}^{(0)}, \boldsymbol{u}_i^{(0)}, \boldsymbol{v}_j^{(0)}, \mathbf{W}_i^{(0)}$
2: **repeatfor** $k = 1, \ldots$
3: $\left[\boldsymbol{x}^{(k)}, \bar{\boldsymbol{x}}^{(k)}, \boldsymbol{u}_i^{(k)}, \boldsymbol{v}_j^{(k)}\right] =$ Algorithm 1 $(\cdots)$
4: $\forall i \in \{1, \ldots, d\}$, **update** $\mathbf{W}_i^{(k)}$
5: **until convergence**
6: **output** $\boldsymbol{x}^{(k)}$

2.5 Re-weighted ℓ_1 minimization

In order to achieve sparsity by analysis of the solution in the ℓ_0 sense, consecutive re-weighted ℓ_1 problems set in (4) are solved in the SARA approach proposed in Carrillo et al. (2012, 2013). We concisely re-explain the re-weighting procedure here for the sake of completeness. In this context, at each iteration indexed by k, a re-weighted ℓ_1 minimization problem associated with the weighting matrix $\mathbf{W}^{(k-1)}$ is solved using the PPD algorithmic structure described in Algorithm 1. The primal and dual variables involved in PPD are initialized from the solutions of the previous weighted ℓ_1 minimization task. Once PPD converges, the weighting matrix $\mathbf{W}^{(k)} = [\mathbf{W}_1^{(k)}, \ldots, \mathbf{W}_b^{(k)}]$ is updated from the previous estimate of the primal variable $\boldsymbol{x}^{(k-1)}$ as follows:

$$\mathcal{D}_e\left(\mathbf{W}_i^{(k)}\right) = \frac{\omega^k}{\omega^k + \alpha_i^{(k)}\left(\left|\boldsymbol{\Psi}_i^\dagger\boldsymbol{x}^{(k-1)}\right|\right)_e}, \tag{11}$$

with the operator $\mathcal{D}_e$ denoting the diagonal element e. The parameter ω is set such that $0 < \omega < 1$, ensuring the decrease of the weights and $\alpha_i^{(k)}$ is basis-dependent and is given by $\alpha_i^{(k)} = 1/\max_e(|\boldsymbol{\Psi}_i^\dagger\boldsymbol{x}^{(k-1)}|)_e$, resulting in scale-free weights. Consequently, the weights are in the interval $[\frac{\omega^k}{2}\ \ 1]$ that tends to $[0\ 1]$ asymptotically. Note that, for each $i \in \{1, \ldots, b\}$, the initial weighting matrix $\mathbf{W}_i^{(0)}$ is the identity matrix. Given this definition, at each re-weighting step, the weights are decreased in such a way that significant analysis coefficients–corresponding to true signal–are strongly down-weighted. After several re-weights, their associated weights tend to zero. By doing so, only small-valued analysis coefficients–typically corresponding to noise–remain highly penalized by the ℓ_1 norm (i.e. their associated weights are close to 1). This weighting scheme is in line with the proposed scheme in Carrillo et al. (2012). The iterative procedure, shown in Algorithm 2, stops when the relative variation between two consecutive estimates is within a bound ϱ where $0 < \varrho < 1$ or the maximum number of iterations is reached.

3 ADAPTIVE PPD FOR REAL IMAGING

Highly sensitive RI data from the new-generation arrays present prominent errors induced by the standard self-calibration, that is an iterative loop alternating between DIE calibration steps and CLEAN imaging steps. DIE modelling errors and lack of DDE calibration yield sky-dependent and correlated errors in the calibrated data. The effect of these errors is reflected in a limited dynamic range of the final recovered radio image. Joint DDE calibration and imaging as proposed in Repetti et al. (2017) would alleviate this effect drastically. However, this is out of the scope of the present article. Herein, the aim is imaging Cyg A from data calibrated with the standard RI pipelines. In this context, we assume that calibration errors share a common scale for each data snapshot, i.e. data aggregated over a short time interval. The blocks in (4) are therefore defined per snapshot and the associated ℓ_2 bounds $\{\epsilon_j\}_{j\in\{1,\ldots,d\}}$ will be set to account not only for the thermal noise but also for the mismodelling of $\boldsymbol{\Phi}$ induced by calibration errors. The level of calibration

errors per data block being a priori unknown, the bounds have to be estimated during image reconstruction. Note that when calibration errors are imperceptible (i.e. buried in the thermal noise) the bounds on the data fidelity terms are fixed with respect to the statistics of the thermal noise (Carrillo et al. 2012; Onose et al. 2016).

With the aim of posing the minimization problem (4) with the most appropriate ℓ_2 constraints, we propose a strategy to adjust adaptively the ℓ_2 bounds on the data fidelity terms during the iterations of the PPD algorithm. The adaptive procedure described below is incorporated in PPD through Steps 6–10 of Algorithm 1 (see modifications coloured in red). Technically, the original ℓ_2 bounds $\{\epsilon_j\}_{j\in\{1,\dots,d\}}$ become iteration-dependent $\{\epsilon_j^{(t)}\}_{j\in\{1,\dots,d\}}$, t being the iteration's index. At each iteration, $\epsilon_j^{(t)}$ is updated as a weighted mean of the current ℓ_2 bound $\epsilon_j^{(t-1)}$ and the ℓ_2 norm of the associated residual data $\mu_j^{(t)} = \|\boldsymbol{y}_j - \boldsymbol{\Phi}_j \boldsymbol{x}^{(t-1)}\|_2$ (Step 8 in Algorithm 1). These updates are performed when the following conditions are met. (i) The estimate of the sky saturates, i.e. the relative variation between two consecutive estimates $\sigma^{(t-1)}$ (calculated in Step 18) is below a fixed value γ_1. (ii) A minimum number of iterations is performed between two consecutive updates of the ℓ_2 bound. (iii) The relative difference between the current estimate of the ℓ_2 bound $\epsilon_j^{(t-1)}$ and the ℓ_2 norm of the corresponding residual data $\mu_j^{(t)}$ is above a certain bound γ_2, where $0 < \gamma_2 < 1$. If the data block does not satisfy its ℓ_2 constraint defined by $\epsilon_j^{(t-1)}$, the latter is assumed underestimated and is increased. Otherwise, it is considered overestimated and is therefore decreased. These conditions are checked at each iteration independently for all the data blocks. Conceptually, the update of the ℓ_2 constraints redefines the minimization problem posed in (4). In this context, conditions (i) and (ii) are set to avoid the early modification of the posed minimization problem, thus ensuring the stability of the strategy.

To initialize the bounds, we perform imaging with the Non-Negative Least Squares (NNLS) algorithm. For each data block indexed by $j \in \{1, \dots, d\}$, we first compute $\tilde{\boldsymbol{x}}_j^{\mathrm{NNLS}}$ that is given by

$$\tilde{\boldsymbol{x}}_j^{\mathrm{NNLS}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \parallel \boldsymbol{y}_j - \boldsymbol{\Phi}_j \boldsymbol{x} \parallel_2^2, \ \text{s.t.}\, \boldsymbol{x} \geq \boldsymbol{0}. \tag{12}$$

We then set the ℓ_2 bound $\epsilon_j^{(0)}$ to $\mu_j^{\mathrm{NNLS}} = \|\boldsymbol{y}_j - \boldsymbol{\Phi}_j \tilde{\boldsymbol{x}}_j^{\mathrm{NNLS}}\|_2$. Since only positivity is imposed, the NNLS minimization problem is under-regularized and the model image tends to over-fit the noisy data. As a consequence, the bounds $\{\epsilon_j^{(0)}\}_{j\in\{1,\dots,d\}}$ tend to be highly under-estimated. Given this initialization, in adaptive PPD, the bounds are adaptively increased while enforcing sparsity. The saturation of the estimate of the solution and consequently the estimates of the bounds are highly correlated with the soft-thresholding parameter κ, inducing sparsity (see Step 13 of Algorithm 1). In fact, when κ is chosen too small, the estimate of the solution converges rapidly in ℓ_2 balls whose bounds are very close to $\{\mu_j^{\mathrm{NNLS}}\}_{j\in\{1,\dots,d\}}$. In this case, the solution is under-regularized and noisy. Whereas, when κ is chosen too high, the estimate of the solution converges in ℓ_2 balls whose bounds are significantly higher than the noise level. In this case, the solution is over-regularized and too sparse. In Onose et al. (2016), the scale-free parameter κ is advised to be set within the interval limited by 10^{-5} and 10^{-3} that is also in line with Carrillo et al. (2014). Though this range remains relevant for adaptive PPD, the algorithm is more sensitive to the choice of the soft-thresholding parameter due to the estimation of the ℓ_2 bounds. For the data imaged herein, a value of order 10^{-5} is found to yield good results.

As detailed above, the resulting algorithmic structure, dubbed adaptive PPD, is very similar to PPD except for its additional feature, which is the ℓ_2 bounds adjustments. An overview of the variables and parameters associated with the adaptive procedure is provided in Appendix A. Formally, PPD is solving the minimization problem set in (4) with well-defined ℓ_2 constraints, whereas adaptive PPD is solving consecutive minimization tasks each corresponding to a different set of ℓ_2 bounds. In fact, if the conditions (i), (ii) and (iii) are met for at least one data block, the adjustment of its associated ℓ_2 bound is performed, hence a new minimization problem is posed and solved, with all variables involved initialized from the last estimates of the previous minimization task. The adaptive PPD algorithm converges when the estimate of the sky saturates and all the ℓ_2 constraints are satisfied. In this case, the weighting matrix involved in the sparsity prior is updated and a new iteration in Algorithm 2 is performed, where a re-weighted ℓ_1 minimization task is solved with adaptive PPD given the current estimates of the ℓ_2 bounds.

4 CYG A IMAGING WITH ADAPTIVE PPD

VLA data sets investigated herein consist of aggregated data acquired with the four configurations of the instrument. The resulting Fourier sampling of the combined data is highly non-uniform. In this case, the preconditioning strategy in the adaptive PPD algorithm is highly effective. Furthermore, since the observations were taken on four different days over a span of over 1 yr, the noise statistics and calibration errors are not consistent over the whole data set. Consequently, assigning different bounds on the data blocks in comparison with assigning one global bound on all the data is crucial. In this section, we present the maps of Cyg A at two frequencies imaged with adaptive PPD. We show the efficiency of the proposed algorithmic structure in recovering superior representations of the radio sky in comparison with the conventional approach MS-CLEAN (Cornwell 2008).

4.1 VLA observations

The data under scrutiny are part of wide-band VLA observations of the well-studied radio galaxy Cyg A within the frequency range of 2–18 GHz, performed over 2 yr (2015–2016). The data sets correspond to observations at X band (8–12 GHz) centred at the frequency 8.422 GHz and C band (4–8 GHz) centred at the frequency 6.678 GHz, each over a spectral window of 128 MHz and with a spectral resolution of 2 MHz. The phase centre is given by RA = $19^{\mathrm{h}}\ 59^{\mathrm{m}}\ 28\overset{\mathrm{s}}{.}356$ ($J2000$) and Dec. = $+40^{\circ}\ 44'\ 2\overset{''}{.}07$. All four configurations (A, B, C, and D) of the VLA have been employed. Their respective total integration times associated with the two data sets are displayed in Table 1. In both observations, the initial time averaging is of 2 s. Decimation of the data sizes is performed via time and frequency averaging over 10 s and 8 MHz. The data sets processed herein are of sizes 2×10^6 and 1.3×10^6 for X band and C band, respectively. They have been carefully

Table 1. Total integration times for the data sets at the two frequencies.

Array configuration	A	B	C	D
8.42 GHz	11.38 h	3 h	1.63 h	0.80 h
6.67 GHz	6.11 h	2.01 h	1.63 h	0.58 h

calibrated using well-established techniques in AIPS, consisting of iterative self-calibration alternating between DIE calibration steps and Cotton–Schwab CLEAN imaging steps (Schwab & Cotton 1983). No DDE calibration is performed, hence the dynamic range on the recovered maps is constrained by the subsequent artefacts rather than the thermal noise.

4.2 Imaging quality assessment

To assess the quality of the reconstructions, we perform visual inspections of the obtained images. These are the estimated model image $\tilde{\boldsymbol{x}}$ and the residual image $\boldsymbol{r} = \beta\boldsymbol{\Phi}^\dagger(\boldsymbol{y} - \boldsymbol{\Phi}\tilde{\boldsymbol{x}})$, where β is a normalization factor.[3] In the context of imaging with the CLEAN-based technique MS-CLEAN, we also consider the restored image $\boldsymbol{z} = \tilde{\boldsymbol{x}} * \boldsymbol{b} + \boldsymbol{r}$, that is the estimated model image convolved with the so-called CLEAN beam $\boldsymbol{b}$, typically a Gaussian fitted to the PSF's primary lobe, and to which the residual image $\boldsymbol{r}$ is added. Convolving the estimated model image with the CLEAN beam is standard in CLEAN imaging. The latter assumes that the sky is made of point sources. Hence, the obtained model image consists of gridded point sources, many of which can be–indeed, need to be–negative. This translates in overemphasizing of the high spatial frequency content of the recovered model image. Such a model of the radio sky is physically unreasonable. Therefore, the standard recourse is to smooth the image. By doing so, a more physical representation of the radio image at the resolution of the instrument is obtained. This is not required for compressive sensing-based approaches. Thanks to the use of both more complex and physical regularizations and explicit data fidelity bounds (see the minimization problem 4); these approaches have been shown to achieve accurate estimates of the ground truth images in synthetic observations and good approximations of the true sky in early real applications (Wiaux et al. 2009a; Wenger et al. 2010; Carrillo et al. 2012; Dabbech et al. 2015; Garsden et al. 2015; Onose et al. 2016; Pratley et al. 2017). Moreover, recent studies have shown that applying a restoring beam on the reconstructed images with this class of methods does not enhance the fidelity to the ground truth image (Chael et al. 2016; Akiyama et al. 2017) as opposed to CLEAN-based methods. Hence, no post-processing convolution by a CLEAN beam or addition of the residual image is recommended. The RI recovered image in the context of compressive sensing-based approaches is the estimated model image.

To quantify the performance of the imaging techniques, we consider the dynamic range metric, which is often adopted in RI imaging and is defined as $\mathrm{DR} = \max_i z_i/\sigma_r$, where σ_r is the standard deviation of the residual image $\boldsymbol{r}$ and $\boldsymbol{z}$ is the restored image as defined in the context of CLEAN imaging. In the computation of the DR values for adaptive PPD, the involved images are obtained as follows. For natural weighting, we compute the residual image $\boldsymbol{r}^{\mathrm{PPD}} = \beta\boldsymbol{\Phi}^\dagger(\boldsymbol{y} - \boldsymbol{\Phi}\tilde{\boldsymbol{x}}^{\mathrm{PPD}})$ and the image $\boldsymbol{z}^{\mathrm{PPD}} = \tilde{\boldsymbol{x}}^{\mathrm{PPD}} * \boldsymbol{b} + \boldsymbol{r}^{\mathrm{PPD}}$. For Briggs weighting, let $\bar{\boldsymbol{y}} = \bar{\boldsymbol{\Theta}}\tilde{\boldsymbol{y}}$ and $\bar{\boldsymbol{\Phi}} = \bar{\boldsymbol{\Theta}}\mathbf{GFZS}$ denote the Briggs-weighted data and their associated measurement operator, where $\bar{\boldsymbol{\Theta}}$ is a diagonal matrix whose elements are the square root of the Briggs weights. We compute the residual image $\bar{\boldsymbol{r}}^{\mathrm{PPD}} = \bar{\beta}\bar{\boldsymbol{\Phi}}^\dagger(\bar{\boldsymbol{y}} - \bar{\boldsymbol{\Phi}}\tilde{\boldsymbol{x}}^{\mathrm{PPD}})$ and the image $\bar{\boldsymbol{z}}^{\mathrm{PPD}} = \tilde{\boldsymbol{x}}^{\mathrm{PPD}} * \bar{\boldsymbol{b}} + \bar{\boldsymbol{r}}^{\mathrm{PPD}}$. The kernels $\boldsymbol{b}$ and $\bar{\boldsymbol{b}}$ are the CLEAN beams associated with natural and Briggs weighting schemes, respectively. However, the DR metric may not reflect accurately the dynamic range in the restored image, since by definition, it is biased by the residual image. In fact, a residual image with a low standard deviation can be induced by false detections in the model image, in particular when positivity is not imposed. This is often the case of CLEAN-based algorithms. Therefore, we report an alternative definition of the dynamic range based on the model image solely, that we call model dynamic range $\mathrm{MDR} = \max_i \tilde{x}_i/\tilde{x}_k$, where $\tilde{x}_k$ is the brightest pixel corresponding to an artefact in the estimated model image.[4] Such a metric is relevant for compressive sensing approaches where the estimated model images are characterized with realistic surface brightness, in particular positive. We do not report MDR for the MS-CLEAN model maps as these exhibit unrealistic features, in particular prominent negative components.

Due to the absence of the ground truth image of the sky, we examine the smoothed versions of the estimated model images at the resolution of the instrument. To compare adaptive PPD with naturally weighted MS-CLEAN, we adopt the images $\tilde{\boldsymbol{z}} = \tilde{\boldsymbol{x}} * \boldsymbol{b}$. Similarly, we adopt the images $\bar{\tilde{\boldsymbol{z}}} = \tilde{\boldsymbol{x}} * \bar{\boldsymbol{b}}$ to compare adaptive PPD with Briggs-weighted MS-CLEAN. We assess the similarity of these two sets of images using the following metric, defined for two signals $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ as $S(\boldsymbol{x}_1, \boldsymbol{x}_2) = 20\log_{10}(\max(\| \boldsymbol{x}_1 \|_2, \| \boldsymbol{x}_2 \|_2)/ \| \boldsymbol{x}_1 - \boldsymbol{x}_2 \|_2)$. We re-emphasize that smoothing the model image obtained with adaptive PPD is not recommended and is performed here only for comparison purposes with MS-CLEAN.

4.3 Imaging results

The performance of adaptive PPD is evaluated in comparison with the standard RI imaging technique MS-CLEAN with two weighting schemes: natural and Briggs. The latter weighting scheme constitutes a compromise between uniform and natural weighting, controlled by a robustness parameter. It is chosen herein over uniform weighting, in the aim to present the optimal reconstructions with MS-CLEAN. On a further note, adaptive PPD imaging is performed in MATLAB and MS-CLEAN imaging is performed using the RI imaging software WSCLEAN (Offringa et al. 2014).

4.3.1 X band

The imaged radio map from the data at a frequency of 8.422 GHz is of size 4096×2048 pixels, with a pixel size $\delta l = 0.04$ arcsec (in both directions). The chosen pixel size is such that $\delta l = 1/5B_X$, where $B_X = \max_{\boldsymbol{u}_\ell \in \{1,\dots,M\}} \|\boldsymbol{u}_\ell\|_2$ is the maximum baseline i.e. the spatial band-limit of the observations. This corresponds to recovering the signal up to 2.5 times the nominal resolution i.e. its recovered spatial bandwidth is $\tilde{B}_X \approx 2.5 \times B_X$. Such a choice of the imaging resolution is conventional in RI imaging.[5] We split the data to 22 blocks of size 10^5 measurements on average, where each block is a single snapshot i.e. data acquired within a time interval over which certain errors (like pointing offsets), can be assumed constant. The number of blocks is chosen to take advantage of the parallelized

[3] Here, we adopt the conventional normalization of the residual image in RI. That is scaling the residual $\boldsymbol{\Phi}^\dagger(\boldsymbol{y} - \boldsymbol{\Phi}\tilde{\boldsymbol{x}})$ by $\beta = 1/\max_i(\boldsymbol{\Phi}^\dagger\boldsymbol{\Phi}\boldsymbol{\delta})_i$, where $\boldsymbol{\delta}$ is an image with value 1 at the phase centre and zero otherwise. By doing so, the PSF defined as $\boldsymbol{h} = \beta\boldsymbol{\Phi}^\dagger\boldsymbol{\Phi}\boldsymbol{\delta}$ has a peak value equal to 1.

[4] The pixel position is determined through the visual inspection of the model image as the one with the highest pixel value and not belonging to the support of the source.

[5] In order to have reasonable results with CLEAN, the PSF needs to be adequately sampled. Therefore, it is common in RI imaging to set the pixel size δl such that $1/5B_X \leq \delta l \leq 1/3B_X$.

structure of adaptive PPD. We perform 70 weighted ℓ_1 minimization tasks using adaptive PPD. Each minimization task stops when the relative variation between two successive estimates of the sky is below 10^{-5}. To ensure higher accuracy of the final solution, the last minimization task stops when the relative variation between two successive estimates of the sky is below 10^{-6}. For MS-CLEAN, we consider imaging with the weighting schemes: natural and Briggs (the robustness parameter is set to $r = -1$). We re-emphasize that in imaging with adaptive PPD, only natural weighting–consisting in whitening the noise–is applied on the measurements in order to reach the optimal sensitivity.

The recovered image of adaptive PPD is displayed in Fig. 1, together with the model and restored images of MS-CLEAN. Three key regions in Cyg A are emphasized: these are the hotspots of the east and west jets (second and third columns), and the inner core of the galaxy (fourth column). When inspecting the model images (rows 1 and 3 of the same figure), one can see that the maps of MS-CLEAN present smooth extended structures since it employs non-delta functions. Though the maps remain non-physical, they are considered more reasonable when compared to the CLEAN algorithm (Högbom 1974; Clark 1980; Schwab & Cotton 1983). Inspection of the restored maps of MS-CLEAN (rows 2 and 4) against the model image of adaptive PPD (fifth row) shows that the latter exhibits more details significantly visible at the hotspots of Cyg A. The validity of this super-resolution is investigated in Section 5.1. Furthermore, both Briggs-weighted MS-CLEAN and adaptive PPD succeed in recovering a faint point source in the inner core of Cyg A, highlighted with a green circle in Fig. 1 (right column, rows 3–5) as opposed to naturally weighted MS-CLEAN (right column, rows 1 and 2). Discussion of this radio source will follow in Section 5.2. Inspection of the residual images displayed in Fig. 2 indicates negative structures at the hotspots positions in the residual image of adaptive PPD. These can be explained by (i) the presence of calibration errors at those positions and (ii) employing CLEAN components in the self-calibration stage that can lead to biased solutions. In fact, errors in CLEAN-like approaches can be absorbed in the model image due to non-positivity. Imaging with the PPD algorithmic structure where positivity is enforced on the estimate of the sky can therefore be in tension with the calibrated data. Adopting PPD in the imaging step during the calibration phase could potentially alleviate these artefacts.

As for the quantitative comparison of the two imaging techniques, we report the similarity of adaptive PPD and naturally weighted MS-CLEAN $S(\tilde{z}^{\text{PPD}}, \tilde{z}^{\text{MS-CLEAN}}) = 32.23$ dB and the similarity of adaptive PPD and Briggs-weighted MS-CLEAN $S(\bar{\tilde{z}}^{\text{PPD}}, \bar{\tilde{z}}^{\text{MS-CLEAN}}) = 32.11$ dB. These values indicate the strong agreement of the recovered low-spatial frequency content with both algorithms, more precisely at the Fourier modes below the spatial band-limit of the observations. The achieved DR values with natural weighting are 6.02×10^3 and 4.26×10^3 for adaptive PPD and naturally weighted MS-CLEAN, respectively. This indicates the higher fidelity achieved by adaptive PPD. On the other hand, the achieved DR values with Briggs weighting are 4.2×10^3 and 7.76×10^3 for adaptive PPD and Briggs-weighted MS-CLEAN, respectively. Note that the latter minimizes the ℓ_2 norm of the residual image $\bar{\boldsymbol{\Phi}}^{\dagger}(\bar{\boldsymbol{y}} - \bar{\boldsymbol{\Phi}}\tilde{\boldsymbol{x}})$, while adaptive PPD minimizes the ℓ_2 norm of the residual image $\boldsymbol{\Phi}^{\dagger}(\boldsymbol{y} - \boldsymbol{\Phi}\tilde{\boldsymbol{x}})$. Conceptually, both methods solve for different imaging problems. Therefore, the higher *DR* achieved by Briggs-weighted MS-CLEAN does not necessarily imply a better performance over adaptive PPD. On a further note, for the sake of comparison, the reported DR values of adaptive PPD are computed using the measurement operators corresponding to the two weighting schemes in WSCLEAN, more precisely, in the computation of the residual images.

The dynamic range on the model image of adaptive PPD is $\text{MDR} = 3.49 \times 10^4$ and is saturated by the DDE modelling errors. The higher value of MDR compared to the DR values of adaptive PPD with both weighting schemes can be justified by two reasons. (i) In the model image of adaptive PPD, the peak value is associated with the central nuclei; the source is super-resolved and its flux is concentrated in few pixels. However, in the adaptive PPD *restored* images (as defined in the context of CLEAN imaging), the source's flux is rather distributed over larger area that is of the size of the adopted CLEAN beam. The peak values in the restored images are instead associated with the hotspots. (ii) The DR value, by definition, may not accurately reflect the level of the noise and errors in the restored image, consisting of the combination of the residual image and the artefacts present in the model image.

4.3.2 C band

The imaged sky at a frequency of 6.678 GHz is of size 3276×1638 pixels, with the pixel size fixed to 0.05 arcsec. We utilize the exact same field of view as in X-band imaging. The spatial bandwidth of the estimated signal is $\tilde{B}_C = 2.5 \times B_C$, where B_C is the maximum baseline. Data are split to 16 blocks with 8×10^4 measurements on average. We perform 60 weighted ℓ_1 minimization tasks using adaptive PPD. Each minimization task stops when the relative variation between two consecutive estimates of the sky gets below 10^{-5}, except for the last minimization task, where the value of this stopping criterion is set to 10^{-6}. MS-CLEAN imaging is performed using the weighting schemes: natural and Briggs (the robustness parameter is set to $r = -1$). Estimated model images of adaptive PPD and MS-CLEAN are displayed in Fig. 3. Superiority of the adaptive PPD reconstructions when compared to those of MS-CLEAN with both weighting schemes in terms of physical surface brightness distribution and high resolution is once again confirmed. The associated residual images are displayed in Fig. 4, where one can see that the adaptive PPD residual image presents the lowest standard deviation. Furthermore, it is less structured when compared to the residual image of naturally weighted MS-CLEAN.

The DR values with natural weighting are 8.04×10^3 and 4.18×10^3 for adaptive PPD and naturally weighted MS-CLEAN, respectively. The DR values with Briggs weighting are 3.45×10^3 and 2.77×10^3 for adaptive PPD and Briggs-weighted MS-CLEAN, respectively. The MDR evaluated on the model image of adaptive PPD is 1.09×10^4. These values indicate higher dynamic range achieved by adaptive PPD, hence higher fidelity. In addition, we report the similarity of the model images obtained with adaptive PPD and MS-CLEAN when smoothed at the instrument's resolution for the two weighting schemes. Similarity of adaptive PPD and naturally weighted MS-CLEAN is $S(\tilde{z}^{\text{PPD}}, \tilde{z}^{\text{MS-CLEAN}}) = 33.51$ dB. Similarity of adaptive PPD and Briggs-weighted MS-CLEAN is $S(\bar{\tilde{z}}^{\text{PPD}}, \bar{\tilde{z}}^{\text{MS-CLEAN}}) = 32.27$ dB. Once again, these results confirm the high similarity of the low spatial frequency content of the recovered images with adaptive PPD and MS-CLEAN.

5 SUPER-RESOLUTION OF CYG A

Both recovered images of Cyg A at bands X and C with adaptive PPD exhibit high-resolution features when compared to the restored maps of MS-CLEAN. In this section, we analyse the super-resolution achieved with adaptive PPD by referring to higher resolution observations of Cyg A. We also confirm the detection of a

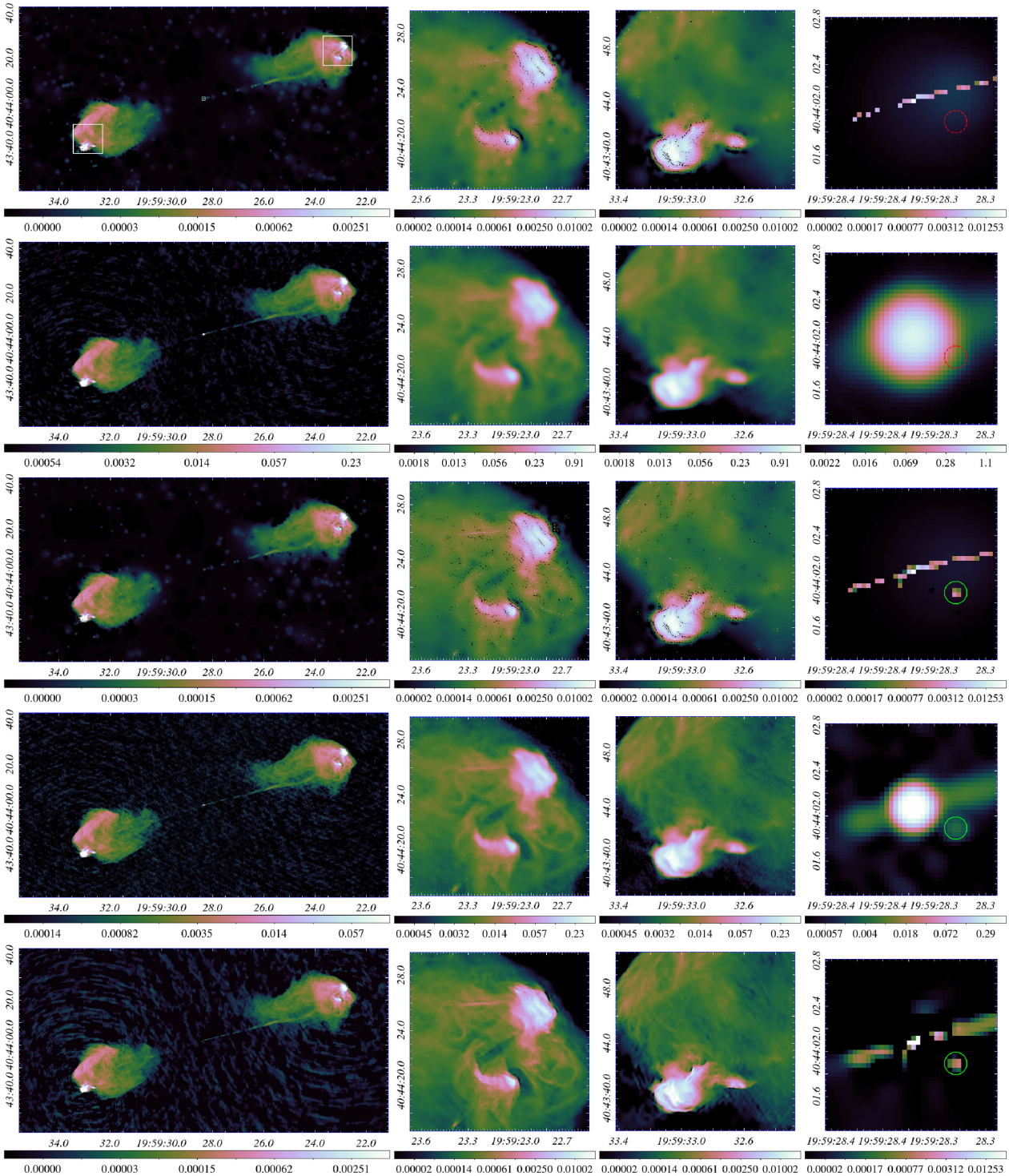

Figure 1. X band: recovered images at 2.5 times the resolution of the instrument. From top to bottom, estimated model and restored images of naturally weighted MS-CLEAN (resp. rows 1 and 2), estimated model and restored images of Briggs-weighted MS-CLEAN (resp. rows 3 and 4) and model image of adaptive PPD (fifth row). The full images are displayed in $\log_{10}$ scale (first column) as well as zooms on the three brightest regions: east jet's hotspot (second column), west jet's hotspot (third column) and the inner core of the Cyg A galaxy (fourth column). The zoomed regions are highlighted with white boxes in the model image of naturally weighted MS-CLEAN (top row, left column). The surface brightness of the restored image obtained with naturally weighted MS-CLEAN (second row) is in units of Jy beam^{-1}, the naturally weighted beam is of size 0.35 arcsec $\times$ 0.35 arcsec and its flux is 90.43 Jy. The surface brightness of the restored image obtained with Briggs-weighted MS-CLEAN (fourth row) is also in units of Jy beam^{-1}, the Briggs-weighted beam is of size 0.18 arcsec $\times$ 0.18 arcsec and its flux is 22.95 Jy. The surface brightness of the model images (rows 1, 3, and 5) is in units of Jy pixel^{-1}, the pixel size being 0.04 arcsec in both directions. Note that the black dots in the hotspots recovered in the model images of MS-CLEAN correspond to important negative pixels. Naturally, these are non-physical components for an intensity map. Therefore, astronomers utilize instead the restored maps, where the prominent negative components disappear thanks to the blurring of the model image with the CLEAN beam.

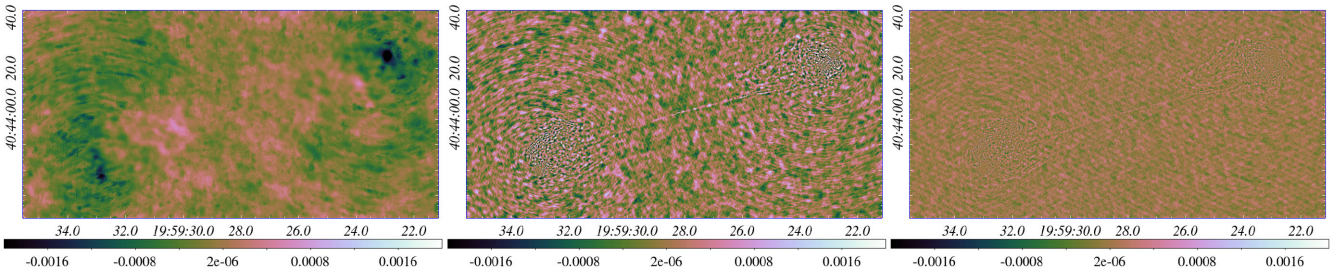

Figure 2. X band: residual images. From left to right: residual image of adaptive PPD with natural weighting ($\sigma_{\boldsymbol{r}} = 2.26 \times 10^{-4}$), MS-CLEAN with natural weighting ($\sigma_{\boldsymbol{r}} = 3.70 \times 10^{-4}$), MS-CLEAN with Briggs weighting ($\sigma_{\tilde{\boldsymbol{r}}} = 1.77 \times 10^{-4}$). The lower value of the standard deviation of the residual image obtained with PPD, compared to that obtained with naturally weighted MS-CLEAN confirms the higher fidelity to data of PPD's estimated model image.

secondary black hole in the inner core of Cyg A, reported in Perley et al. (2017), and super-resolved in the maps produced by adaptive PPD.

5.1 Analysis of the super-resolution with adaptive PPD

To judge the veracity of the super-resolution capabilities of adaptive PPD, we compare its results to maps of Cyg A obtained from higher-frequency observations, where the nominal resolution is naturally higher, hence super-resolution is not required. More precisely, we choose VLA observations at Ku band (12–18 GHz), where the nominal resolution is two to three times that of the X and C band data. Their associated maps are restored images obtained with the Cotton-Schwab CLEAN (Schwab & Cotton 1983). Since the emission mechanism in Cyg A is known to be synchrotron, the radiation spectrum is very broad, spanning over orders of magnitudes (Carilli et al. 1991). No sharp spectral features are expected within a frequency band of the same order of magnitude. Therefore, the choice of Ku band images as a reference is relevant.

For the imaged data at X band, we utilize a restored map of Cyg A at a frequency of 17.324 GHz as a reference. These data present a maximum baseline B_1^{ref}, that is $B_1^{\text{ref}} \approx 2 \times B_X$, and the spatial bandwidth of its imaged map is $\tilde{B}_1^{\text{ref}} = 4 \times B_X$. The X band data, initially imaged at the spatial bandwidth $2.5 \times B_X$, are further imaged at the exact same spatial bandwidth as the reference image $\tilde{B}_1^{\text{ref}}$ using both Briggs-weighted MS-CLEAN and adaptive PPD. In Fig. 5, zooms on the hotspots are displayed for the reference image, the model image of adaptive PPD and the restored image of Briggs-weighted MS-CLEAN. All three images are characterized with a pixel size $\delta l = 0.025$ arcsec. The surface brightness of PPD model images are in units of Jy pixel^{-1} and that of MS-CLEAN and the reference maps are in units of Jy beam^{-1}. Note that the displayed integrated flux is preserved on the three maps. The inspection of the zooms on both hotspots indicates the high similarity of the recovered hotspots in the model image of adaptive PPD and the CLEAN restored image of Cyg A at the frequency 17.324 GHz. A further examination of adaptive PPD's consistency with respect to the choice of the imaging resolution is examined through the image recovery at three different resolutions. These correspond to the pixel sizes $\delta l \in \{0.08,\ 0.04,\ 0.025\,\text{arcsec}\}$, shown in Fig. 5 as embedded animations[6] cycling through the imaged hotspots at the different resolutions. The high-resolution features are consistent over the different resolutions with a noticeable improvement when increasing the imaged spatial bandwidth, in particular in terms of the pixelization at the edges of the hotspots' brightest structures. Thus, super-resolved maps up to four times the nominal resolution can be obtained with no apparent degradation of the imaging quality despite the increase of the number of the unknowns in the imaging problem. Note that when running MS-CLEAN on a resolution larger than the nominal one, naturally a super-resolved model image is obtained, where the negative components are less prominent. However, as explained earlier, the super-resolved structures of the MS-CLEAN model image remain non-physical. Moreover, the resolution of the corresponding restored image is limited by the shape of the CLEAN beam, dictated by the effective Fourier sampling. This effect is illustrated in Fig. 5, where one can see that (i) the hotspots recovered in the restored image of Briggs-weighted MS-CLEAN are smooth when compared against the reconstruction of adaptive PPD, (ii) no super-resolution is noticed when examining the hotspots at the three different resolutions.

Similar investigation is carried out for the obtained images from the data at C band. These are cross-checked with a restored image of Cyg A at a frequency of 14.252 GHz. The maximum baseline of the observations is $B_2^{\text{ref}} \approx 2.13 \times B_C$ and the imaged map's spatial bandwidth is $\tilde{B}_2^{\text{ref}} \approx 3.5 \times B_C$ corresponding to a pixel size $\delta l = 0.035$ arcsec. We perform imaging with adaptive PPD and Briggs-weighted MS-CLEAN at the same spatial bandwidth $\tilde{B}_2^{\text{ref}}$ (i.e. same resolution). Zooms on the hotspots of the reference map, the adaptive PPD model image and the Briggs-weighted MS-CLEAN restored image are shown in Fig. 6. The surface brightness of PPD model images are in units of Jy pixel^{-1} and that of MS-CLEAN and the reference maps are in units of Jy beam^{-1}. Once again, when inspecting visually the hotspots, it is clear that the obtained details with adaptive PPD are physical. Super-resolution recovery of adaptive PPD is again confirmed. Moreover, inspecting the obtained maps with adaptive PPD at the different resolutions, characterized with pixel sizes $\delta l \in \{0.08, 0.05, 0.035\,\text{arcsec}\}$, demonstrates the consistency of the high-resolution features of the algorithm. These images are shown in Fig. 6 as embedded animations cycling through the imaged hotspots at the different resolutions.

These super-resolved maps of adaptive PPD are the result of the SARA priors enforced on the estimate of the sky, which are the positivity and the re-weighted average sparsity in a redundant dictionary. In particular, positivity seems to have a significant impact on the spatial frequencies above the maximum baseline of the observations. For a better understanding of this prior's contribution, we note that the reconstructed image of the radio sky is a sampling of a spatially band-limited version of the true sky. Moreover, if exact, it can be expressed as the sampled convolution of the true positive sky with a *sinc* function. Thus, conceptually, it can take negative values. Therefore, formally, the positivity constraint comes in tension with the nature of the true samples of the spatially band-limited sky. Yet,

[6] The animation is only supported when the PDF file is opened using Adobe Acrobat Reader, https://get.adobe.com/reader/.

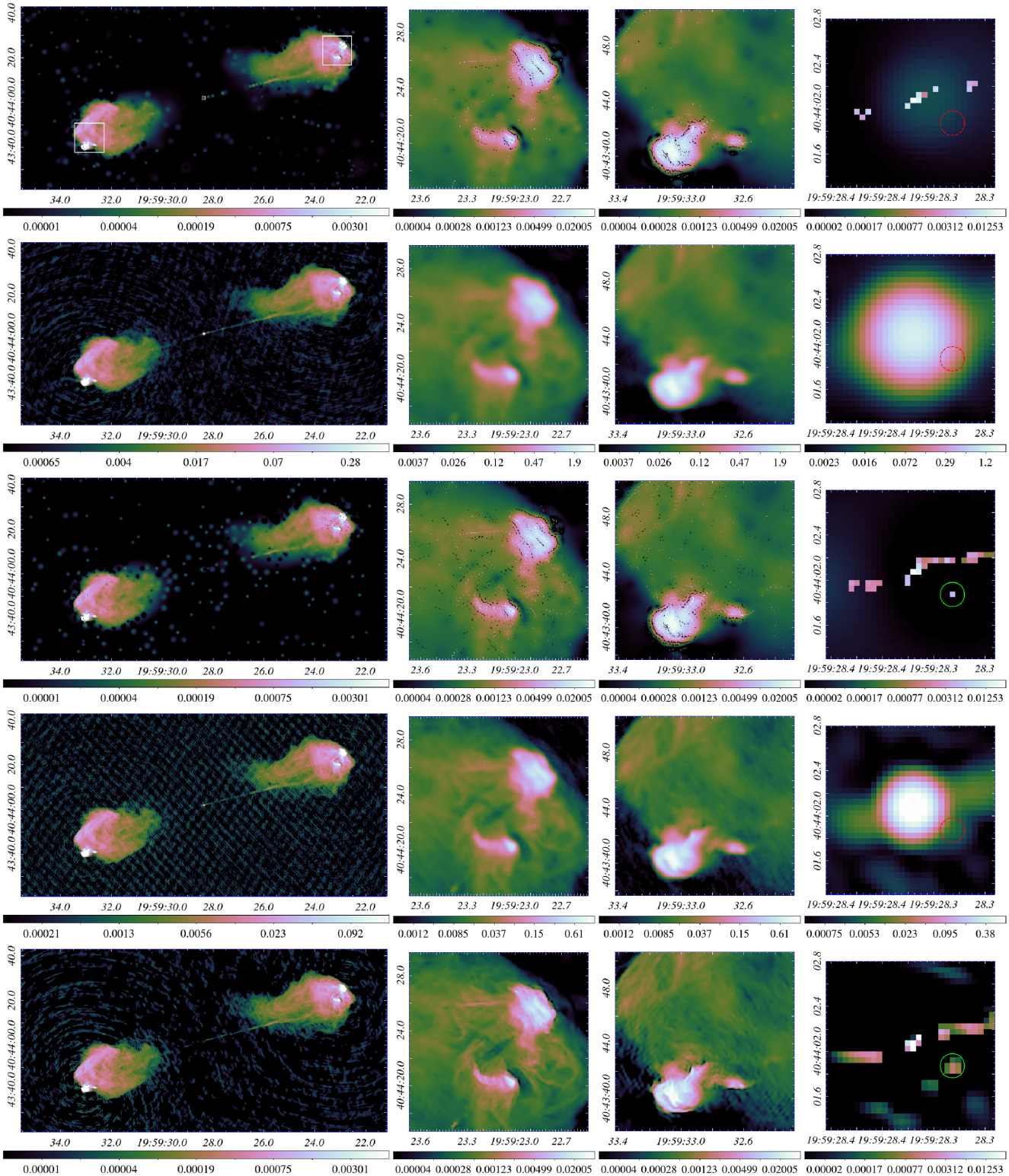

Figure 3. C band: recovered images at 2.5 times the resolution of the instrument. From top to bottom, estimated model and restored images of naturally weighted MS-CLEAN (resp. rows 1 and 2), estimated model and restored images of Briggs-weighted MS-CLEAN (resp. rows 3 and 4) and model image of adaptive PPD (fifth row). The full images are displayed in $\log_{10}$ scale (first column) as well as zooms on the three brightest regions: east jet's hotspot (second column), west jet's hotspot (third column), and the inner core of the Cyg A galaxy (fourth column). The zoomed regions are highlighted with white boxes in the model image of naturally weighted MS-CLEAN (top row, left column). The surface brightness of the restored image obtained with naturally weighted MS-CLEAN (second row) is in units of Jy beam^{-1}, the naturally weighted beam is of size 0.45 arcsec × 0.45 arcsec and its flux is 93.65 Jy. The surface brightness of the restored image obtained with Briggs-weighted MS-CLEAN (fourth row) is also in units of Jy beam^{-1}, the Briggs-weighted beam is of size 0.25 arcsec × 0.25 arcsec and its flux is 30.44 Jy. The surface brightness of the model images (rows 1, 3 and 5) is in units of Jy pixel^{-1}, the pixel size being 0.05 arcsec in both directions.

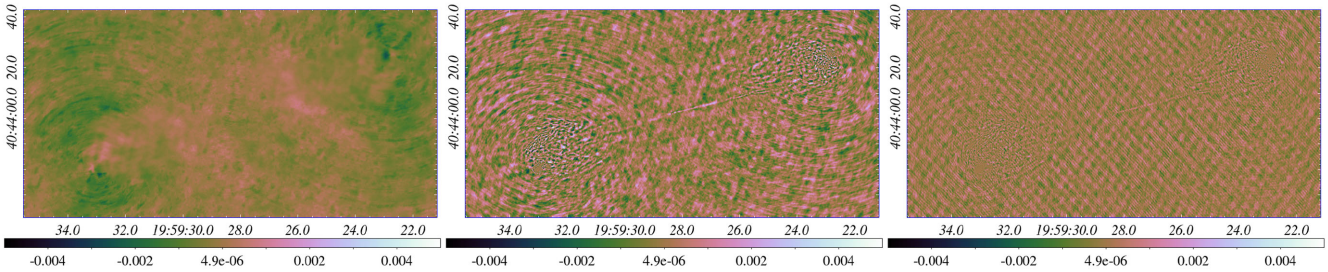

Figure 4. C band: residual images. From left to right: images of adaptive PPD with natural weighting ($\sigma_{\boldsymbol{r}} = 3.65 \times 10^{-4}$), MS-CLEAN with natural weighting ($\sigma_{\boldsymbol{r}} = 7.01 \times 10^{-4}$), MS-CLEAN with Briggs weighting ($\sigma_{\tilde{\boldsymbol{r}}} = 4.60 \times 10^{-4}$). The lower value of the standard deviation of the residual image obtained with PPD, compared to that obtained with naturally weighted MS-CLEAN confirms the higher fidelity to data of PPD's estimated model image.

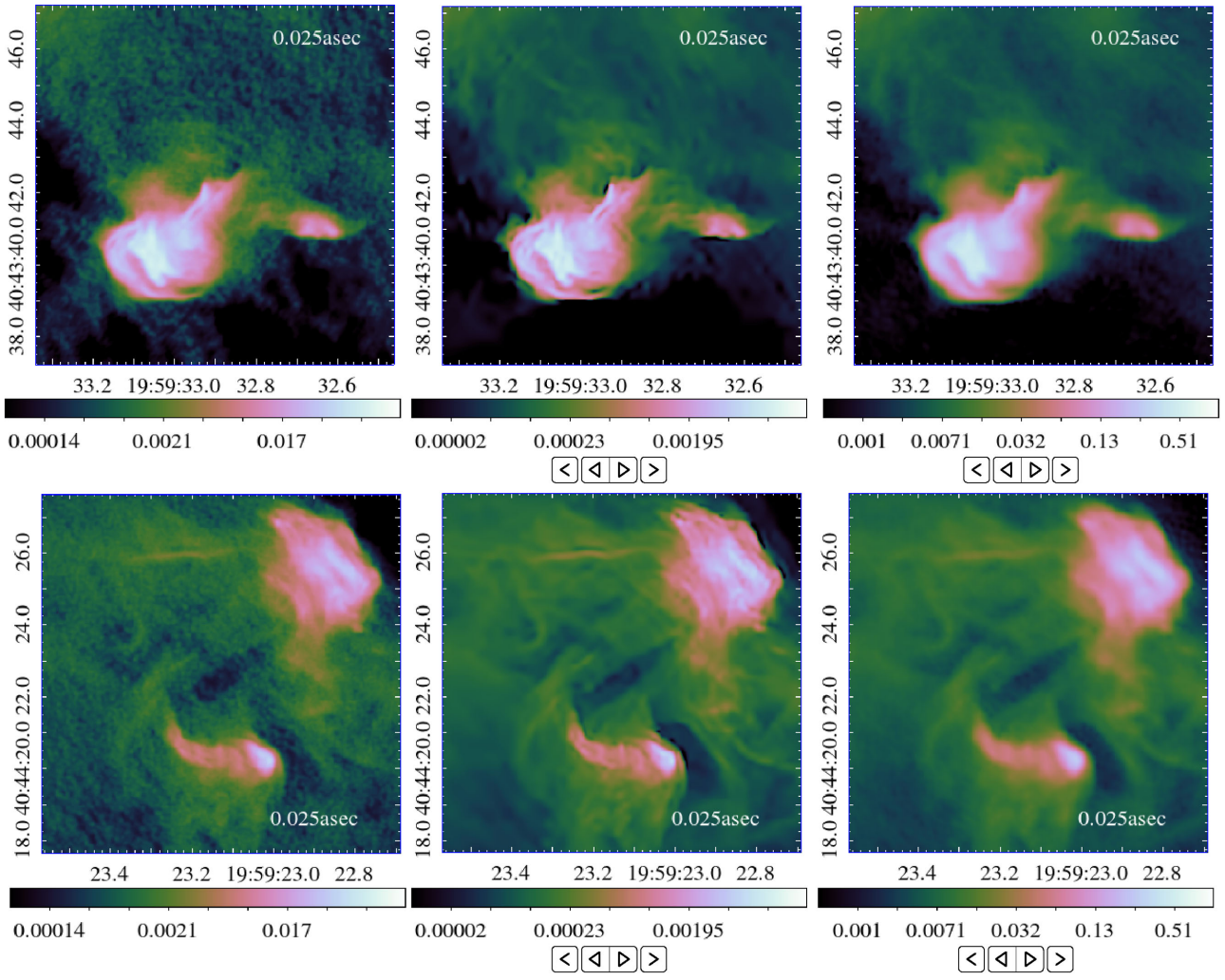

Figure 5. X band: zooms on the hotspots in Cyg A displayed in $\log_{10}$ scale. Top: east hotspot, bottom: west hotspot. From left to right: the reference map at 17.324 GHz obtained with CLEAN, the estimated model image of adaptive PPD, and the restored image of Briggs-weighted MS-CLEAN from the data at X band (8.422 GHz). All images have the same pixel size $\delta l = 0.025$ arcsec. The figure also contains embedded animation of the hotspots imaged with adaptive PPD and Briggs-weighted MS-CLEAN at three different resolutions corresponding to $\delta l \in \{0.08, 0.04, 0.025\,\text{arcsec}\}$. The surface brightness of Briggs-weighted MS-CLEAN and the reference map are in units of Jy beam^{-1}. The surface brightness of PPD's model images are in units of Jy pixel^{-1}. Note that, for each resolution, the unit of the surface brightness is different as it is a function of the pixel. The displayed integrated flux is preserved on all the maps. One can notice improved details with increased resolution, in particular at the edges of the brightest structures of the hotspots. The animations are only supported when the PDF file is opened using Adobe Acrobat Reader.

this tension can be alleviated by choosing to image the radio sky at a spatial bandwidth significantly larger than the maximum baseline of the observations. In this case, positivity simply acts as a strong prior for the Fourier modes beyond the maximum baseline of the observations, i.e. for super-resolution. As shown above, super-resolution obtained with adaptive PPD is validated. These findings confirm that the estimated model images using convex optimization techniques are highly reliable, thus no post-processing, such as in-

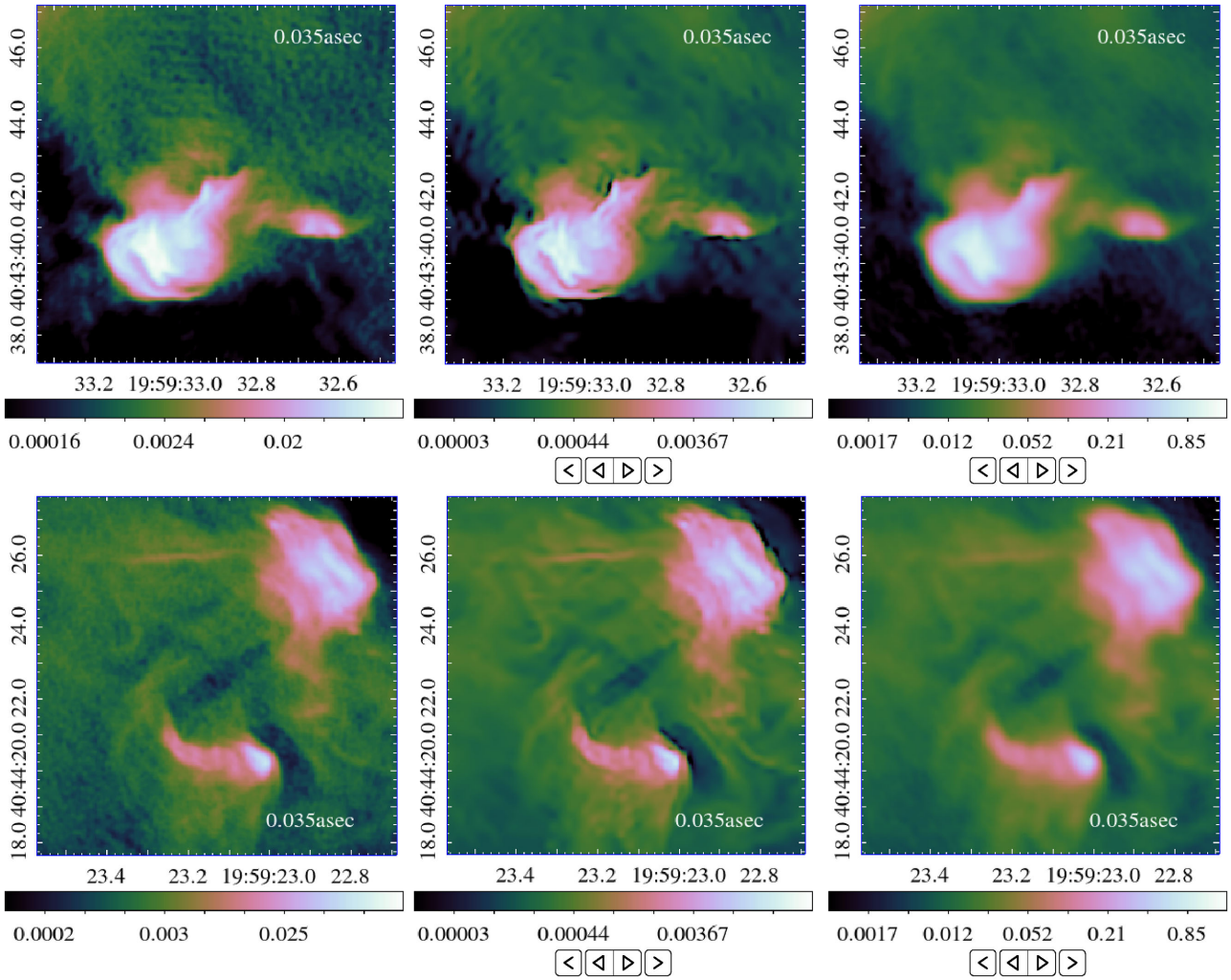

Figure 6. C band: zooms on the hotspots in Cyg A displayed in $\log_{10}$ scale. Top: east hotspot, bottom: west hotspot. From left to right: the reference map at 14.252 GHz obtained with CLEAN, the estimated model image of adaptive PPD and the restored image of Briggs-weighted MS-CLEAN from the data at C band (6.678 GHz). All images have the same pixel size $\delta l = 0.035$ arcsec. The surface brightness of Briggs-weighted MS-CLEAN and the reference map are in units of Jy beam^{-1}. The surface brightness of PPD's model images are in units of Jy pixel^{-1}. Note that, for each resolution, the unit of the surface brightness is different as it is a function of the pixel. The displayed integrated flux is preserved on all the maps. The figure also contains embedded animations of the hotspots imaged with adaptive PPD and Briggs-weighted MS-CLEAN at three different resolutions, corresponding to $\delta l \in \{0.08, 0.05, 0.035$ arcsec$\}$. One can notice improved details with increased resolution, in particular at the edges of the brightest structures of the hotspots. The animations are only supported when the PDF file is opened using Adobe Acrobat Reader.

troducing a blur reflecting the instrument's resolution, is required. On a further note, the *w*-modulation, originating from the third dimension of the baseline, has recently been shown to yield super-resolution (Dabbech et al. 2017). Yet, this is not the case here, as the probed field of view is narrow. The *w*-modulation, being negligible, is not considered in the imaging problem.

5.2 The story of a secondary black hole (candidate)

Perley et al. (2017) report the serendipitous discovery of a luminous radio transient in the inner core of Cyg A, just 460 pc offset from the supermassive black hole in the galaxy. The transient, dubbed Cyg A-2, is well detected at 8.5 GHz using CLEAN, and is interpreted as a secondary black hole. We confirm the findings of Perley et al. (2017) when imaging Cyg A from observations at a frequency of 8.422 GHz, at 2.5 times the nominal resolution (corresponding to a pixel size of $\delta l = 0.04$ arcsec). In Fig. 1, fourth column, zooms on the inner core of Cyg A are displayed. The source's location is highlighted with a circle whose centre is at the position given by RA $= 19^{\mathrm{h}}59^{\mathrm{m}}28\overset{\mathrm{s}}{.}322$ (J2000) and Dec. $= +40^{\circ}\,44'\,1\overset{''}{.}89$ and radius of size 0.1 arcsec. The source is highlighted with a green circle when detected and a red dashed circle otherwise. One can see that Cyg A-2 is well detected with adaptive PPD and Briggs-weighted MS-CLEAN, whereas naturally weighted MS-CLEAN fails to do so. Though visible, the source is blurred in the restored image of Briggs-weighted MS-CLEAN. The flux of Cyg A-2, calculated directly from the model images (over the highlighted physical region), is about 5 mJy with adaptive PPD and 4.9 mJy with Briggs-weighted MS-CLEAN. The source is also well detected when imaged at four times the nominal resolution (corresponding to a pixel size $\delta l = 0.025$ arcsec) as shown in Fig. 7 (left-hand panel), where its angular scale is preserved and its estimated flux is about 4.3 mJy.

More interestingly, the source is highly resolved when imaged with adaptive PPD from the observations at a frequency of

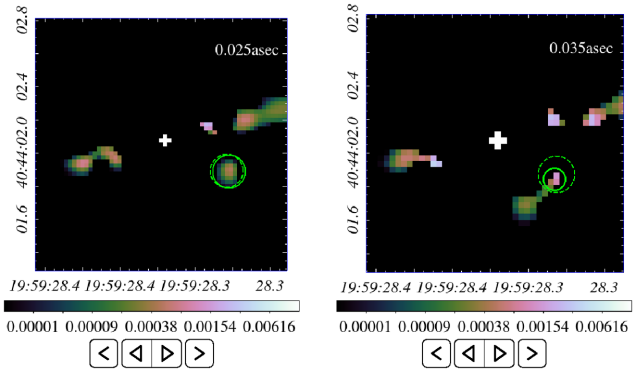

Figure 7. Cyg A-2 displayed in $\log_{10}$ scale. Left: X band observations imaged with PPD at four times the nominal resolution ($\delta l = 0.025$ arcsec). Right: C band observations imaged with PPD at 3.5 times the nominal resolution ($\delta l = 0.035$ arcsec). Cyg A-2 is highlighted with a green circle. The source's angular scales from the maps at 2.5 times the nominal resolutions are highlighted with green dashed circles. These maps are embedded as animations in the figure. The animations are only supported when the PDF file is opened using Adobe Acrobat Reader.

6.678 GHz at 2.5 times the nominal resolution (corresponding to a pixel size $\delta l = 0.05$ arcsec), as shown in Fig. 3 (fourth column). The source's location is highlighted with a circle whose centre is at the position given by RA $= 19^{\rm h}\,59^{\rm m}\,28\overset{\rm s}{.}324$ ($J2000$) and Dec. $= +40^\circ\,44'\,1\overset{''}{.}88$ and radius of size 0.11 arcsec. The source is highlighted with a green circle when detected and a red dashed circle otherwise. Although detected in the model image of MS-CLEAN with Briggs weighting at a single pixel, in the restored image, the source is completely buried within the beam of the primary nuclei of Cyg A (see Fig. 3, fourth column, fourth row). As for MS-CLEAN with natural weighting, here again it fails completely to detect the radio transient. The estimated flux of Cyg A-2 is about 4.6 mJy with both adaptive PPD and Briggs-weighted MS-CLEAN (over the highlighted physical region). These results are in agreement with the findings of Perley et al. (2017). The source is also well detected when imaged at 3.5 times the nominal resolution (corresponding to a pixel size $\delta l = 0.035$ arcsec) as shown in Fig. 7 (right-hand panel), where it is further resolved and its estimated flux is about 3.4 mJy. We note the presence of a tail-like structure associated with Cyg A-2 in the PPD image. However, given the faintness of the structure, and the fact that it is not detected in the X-band image, we cannot confidently say whether this faint structure is real, or is a DDE-induced imaging artefact.

6 CONCLUSIONS

In this paper, we developed an adaptive version of the convex PPD algorithmic structure solving the SARA minimization problem for RI imaging in the presence of unknown noise and calibration errors. The algorithm achieves high-resolution high-fidelity imaging of Cyg A from VLA observations. Imaging results confirm the superior quality of the proposed algorithmic structure to standard CLEAN-based techniques. The veracity of the achieved super-resolved reconstructions of Cyg A at X and C bands is verified through higher resolution VLA observations of the radio galaxy at Ku band. These results confirm the reliability of the reconstructed images with the advanced convex optimization algorithms as accurate representations of the radio sky. Our MATLAB code is available online on GitHub, https://github.com/basp-group/Puri-Psi/. Interestingly, the recent discovery of a radio transient in the inner core of Cyg A, revealed at X band, is further confirmed at C band when imaging with adaptive PPD, though the latter observations are characterized with a lower nominal resolution. The radio source is very well detected on the adaptive PPD model images at both frequencies and is super-resolved when compared against the restored images obtained with MS-CLEAN.

ACKNOWLEDGEMENTS

The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. The research of AD, AO, AA, and YW is supported by the UK Engineering and Physical Sciences Research Council (EPSRC, grants EP/M008843/1, EP/M011089/1). The research of OS is supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation.

REFERENCES

Akiyama K. et al., 2017, ApJ, 838, 1
Bhatnagar S., Cornwell T. J., 2004, A&A, 426, 747
Candès E. J., 2006, Int. Congress Math. Madrid, Spain
Candès E. J., Wakin M. B., Boyd S. P., 2008, J. Fourier Anal. Appl., 14, 877
Carilli C. L., Perley R. A., Dreher J. W., Leahy J. P., 1991, ApJ, 383, 554
Carrillo R. E., McEwen J. D., Wiaux Y., 2012, MNRAS, 426, 1223
Carrillo R. E., McEwen J. D., Ville D. V. D., Thiran J.-P., Wiaux Y., 2013, IEEE Sig. Proc. Let., 20, 591
Carrillo R. E., McEwen J. D., Wiaux Y., 2014, MNRAS, 439, 3591
Chael A. A., Johnson M. D., Narayan R., Doeleman S. S., Wardle J. F. C., Bouman K. L., 2016, ApJ, 829, 11
Clark B. G., 1980, A&A, 89, 377
Combettes P. L., Pesquet J.-C., 2011, in Bauschke H. H., Burachik R. S., Combettes P. L., Elser V., Luke D. R., Wolkowicz H., eds, Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer-Verlag, New York, p. 185
Cornwell T. J., 2008, IEEE Sel. Top. Sig. Proc., 2, 793
Cornwell T., Golap K., Bhatnagar S., 2008, IEEE Sel. Top. Sig. Proc., 2, 647
Dabbech A., Mary D., Ferrari C., 2012, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE Kyoto, Japan, p. 3665
Dabbech A., Ferrari C., Mary D., Slezak E., Smirnov O., Kenyon J. S., 2015, A&A, 576, A7
Dabbech A., Wolz L., Pratley L., McEwen J. D., Wiaux Y., 2017, MNRAS, 471, 4300
Garsden H. et al., 2015, A&A, 575, A90
Högbom J. A., 1974, A&A, 15, 417
Li F., Cornwell T. J., de Hoog F., 2011, A&A, 528, A31
Offringa A. R. et al., 2014, MNRAS, 444, 606
Onose A., Carrillo R. E., Repetti A., McEwen J. D., Thiran J.-P., Pesquet J.-C., Wiaux Y., 2016, MNRAS, 462, 4314
Onose A., Dabbech A., Wiaux Y., 2017, MNRAS, 469, 938
Perley D. A., Perley R. A., Dhawan V., Carilli C. L., 2017, ApJ, 841, 117
Pesquet J.-C., Repetti A., 2015, J. Nonlinear Convex Anal., 16, 2353
Pratley L., McEwen J. D., d Avezac M., Carrillo R. E., Onose A., Wiaux Y., 2017, MNRAS, 473, 1038
Repetti A., Birdi J., Dabbech A., Wiaux Y., 2017, MNRAS, 470, 3981
Schwab F. R., Cotton W. D., 1983, AJ, 88, 688
Wakker B. P., Schwarz U. J., 1988, A&A, 200, 312
Wenger S., Magnor M., Pihlström Y., Bhatnagar S., Rau U., 2010, PASP, 122, 1367
Wiaux Y., Jacques L., Puy G., Scaife A. M. M., Vandergheynst P., 2009a, MNRAS, 395, 1733
Wiaux Y., Puy G., Boursier Y., Vandergheynst P., 2009b, MNRAS, 400, 1029

APPENDIX A: OVERVIEW OF THE PARAMETERS SPECIFIC TO ADAPTIVE PPD

An overview of the variables and parameters involved in the adjustment of the ℓ_2 bounds on the data fidelity terms is presented in Tables A1 and A2, respectively.

Table A1. Overview of the variables employed in the adaptive procedure incorporated in Algorithm 1.

$\mu_j^{(t)}$	ℓ_2 norm of the residual corresponding to the data block j at iteration t.
$\epsilon_j^{(t-1)}$	ℓ_2 bound on the data block j imposed at iteration t.
$p_j^{(t-1)}$	Iteration index of the previous update of the ℓ_2 bound of the data block j.
$\sigma^{(t-1)}$	Characterizing the relative variation between two consecutive estimates of the solution at iteration $t-1$.

Table A2. Overview of the parameters involved in the adaptive procedure incorporated in Algorithm 1.

$\gamma_1 \in]01[$	Configurable; the bound on the relative variation between two consecutive estimates of the solution. For the tests herein γ_1 is set to 10^{-4}.
$\gamma_2 \in]01[$	Configurable; the tolerance on the relative difference between the current estimate of the bound imposed on the data block j and the ℓ_2 norm of its associated residual. For the tests herein γ_2 is set to 10^{-3}.
$\gamma_3 \in]01[$	Configurable; characterizing the increment of the ℓ_2 bound with respect to the ℓ_2 norm of the current residual. For the tests herein γ_3 is set to 0.618.
P	Configurable; corresponds to the minimum number of iterations between consecutive updates on each ℓ_2 bound. For the tests herein P is set to 100.

This paper has been typeset from a TEX/LATEX file prepared by the author.